DOCUMENT RESUME

ED 199 052 SE 034 208

AUTHOR	Schoen, Harold L.; And Others
TITLE	The Iowa Problem-Solving Project: Development and Evaluation.
PUB DATE	Aug 80
NOTE	38p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Calculators; *Curriculum Development; Educational Research; Elementary Secondary Education; Instructional Materials; Mathematics Curriculum; *Mathematics Education; *Mathematics Instruction; *Problem Solving; Teaching Methods
IDENTIFIERS	*Iowa Problem Solving Project; *Mathematics Education Research

ABSTRACT

This report discusses some of the activities of the Iowa Problem-Solving Project (IPSP). The document has three main sections: (1) Materials Development, where IPSP teaching materials and test development are discussed; (2) Summative Evaluation of Materials, presenting the procedures, results, and findings of the 1978-79 evaluations; and (3) Further Discussion, with the IPSP approach and results placed in the broader perspective of research on methods of teaching problem solving. The IPSP was created to develop, evaluate, and disseminate materials to improve problem-solving abilities in grades five through eight. The four main assumptions are: (1) pupils need to learn to solve many nontrivial problems of interest to them; (2) specific strategies for solving these problems should be taught; (3) a general framework is useful in teaching problem solving that can be used to organize thinking; and (4) it is possible and pedagogically useful to test problem-solving abilities within the steps of the framework. The material incorporates hand-held calculators. Results of evaluation reported by this document attest to the effectiveness of the IPSP approach, as measured by posttesting and attitude changes of students and teachers. (MP)

ED199052

The Iowa Problem-Solving Project:

Development and Evaluation

by

Harold L. Schoen

University of Iowa

Theresa Oehmke

University of Iowa

John Tarr

University of Northern Iowa

August 1980

Running Head: Iowa Problem Solving

SE 034 208

The purpose of this paper is to report on some of the activities of the Iowa Problem-Solving Project (IPSP). The paper is divided into three main sections: Materials Development, in which the IPSP teaching materials and test development are discussed; Summative Evaluation of the Materials, in which the procedures, results and findings of the 1978-79 IPSP evaluation are presented; Further Discussion, in which the IPSP approach and results are placed in the broader perspective of research on methods of teaching problem solving.

MATERIALS DEVELOPMENT

The Iowa Problem-Solving Project, directed by George Immerzeel, was first funded in 1976 under Public Law 93-380, Title IV, Part C to develop, evaluate, and disseminate materials to improve the mathematical problem-solving abilities of students in grades 5, 6, 7, and 8. Thus, the major problem addressed by IPSP was, "Can materials be produced which improve the mathematical problem-solving abilities of students in grades 5-8?"

Parents, teachers, and school administrators generally agree that one of the most important goals of school mathematics is to develop in each student the ability to solve problems. Although the importance of problem solving in mathematics is well recognized, it is equally recognized that past practices have not been entirely successful. The National Advisory Committee on Mathematical Education (NACOME), discussing applications, reported,

> Inspection of current commercial texts, standardized tests, national assessment items, or state and local mathematical syllabi confirms the disappointing impression that "application" in school mathematics means "word problem." For most

> of these problems, the main task for students is translating the technical jargon of mathematical prose into simpler language and then into suitable symbolic form. Furthermore, subsequent arithmetic and algebraic manipulations inevitably lead to simple closed-form solutions which students quite accurately see applying to few realistic situations. (NACOME, 1975, p. 26)

The relatively poor showing of students attempting to solve word problems was documented in reports of the first National Assessment of Educational Progress (NAEP). In stating implications of NAEP for instruction, Carpenter et al concluded,

> It is most disturbing to entertain the suggestion that many students receive very little opportunity to learn to solve word problems. The assessment results are so poor, however, that we wonder whether this is not the case. A commitment to working and thinking about word problems is needed for teachers and their students. (Carpenter, et al, 1976, p. 392)

This situation had not improved five years later as the results of the second round of NAEP indicate.

> The results of a large number of exercises at the problem-solving level ... clearly demonstrate that <u>many students lack even the most basic problem-solving skills</u>.
>
> Although the specific question of instruction is not addressed by the assessment, we believe that in order to improve students' problem-solving abilities <u>specific attention must be given to teaching problem-solving strategies</u>. (Carpenter, et al, 1980, p. 430)

Although the development of children's problem-solving abilities is a major goal of elementary school mathematics, there is little evidence that a serious attempt was being made to attain this goal (Lester, 1980, p. 287). And yet much is known about how to improve problem solving ability. For example, it is known that just giving many problems of appropriate difficulty to students will likely effect some increment in ability to solve problems (Callahan, 1975, pp. 145-149). The use of interesting problems and teaching specific problem solving strategies are also supported by research (Lester, 1980).

The IPSP approach to teaching problem solving rests on four main assumptions.

1. Students need to spend time learning to solve and then actually solving many nontrivial problems, which are of interest to them.
2. Specific strategies for solving these problems should be taught to the students. One way to do this is to organize the instruction and practice around problem-solving strategies rather than around concepts or algorithms.
3. A general framework is useful in teaching problem solving and that same framework can be used by the students to organize their thinking as they attempt to solve problems. It can also provide a convenient language for discussing problem solving.
4. It is possible and pedagogically useful to test students' problem-solving abilities within the steps of the problem-solving framework. Such testing would reinforce to the students and teachers the importance of the entire range of skills needed in problem solving. Thus, it would encourage thinking through problems, looking back at solutions, etc. as well as finding the correct answers.

To achieve its goal, IPSP developed instructional materials for students and their teachers. The materials incorporated use of hand-held calculators which, in the mid 1970s, were just becoming widely available at low cost. Use of calculators in problem solving had been supported in the NACOME report which stated,

> Availability of a calculator does not remove the necessity of analyzing problem situations to determine appropriate calculations and to interpret correctly the numerical results. ... With de-emphasis on the purely mechanical aspects of arithmetic comes an opportunity to pay close attention to other crucial aspects of the problem-solving process and to treat more genuine problems with the "messy" calculations they inevitably involve. (NACOME, 1975, pp. 42, 43)

The Iowa Problem-Solving Project developed eight instructional modules:

Problem Solving Using the Calculator--Book 1
Problem Solving Using the Calculator Codes--Book 1
Problem Solving Using Guesses
Problem Solving Using Tables
Problem Solving Using the Calculator--Book 2
Problem Solving Using Calculator Codes--Book 2
Problem Solving Using Resources
Problem Solving Using Special Computations

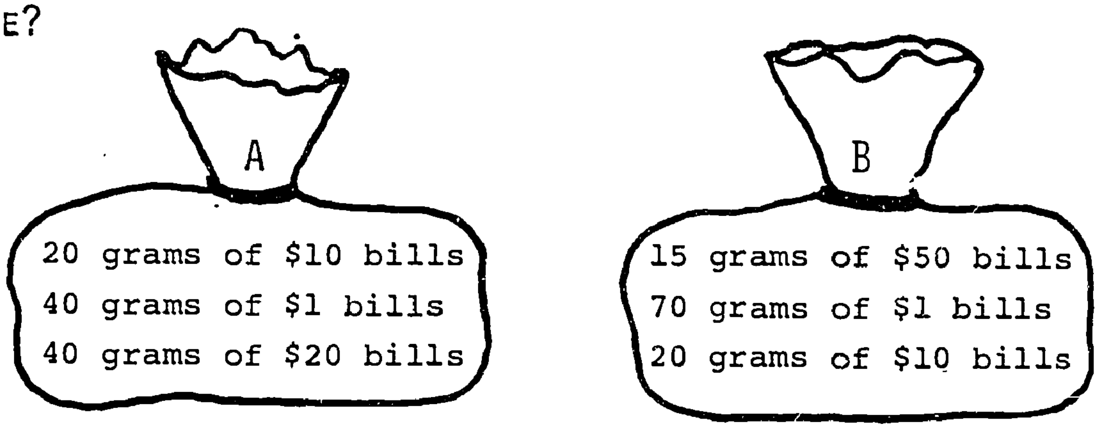
The four listed first were written for grades 5-6; the other four are for grades 7-8. All the materials assume that hand-held calculators are available to the students--one calculator for every two students. During the development phase, the IPSP provided calculators to tryout classrooms where they would not otherwise have been available.

Each module requires about 10 days of class time and consists of three components--a student booklet, a deck of 100 problem cards, and a teacher's guide. The 25-page student booklets address the problem-solving skills used in that module such as using guesses, using tables, using calculator codes, and using resource books. Typically working in partnerships, students spend about five days working from the booklet under the teacher's supervision. After completing the skills booklet, the partnership solves problems they select from the deck of 100 cards written to provide practice in using the strategies developed in the module. The cards are color-coded to indicate the problem difficulties. The problems provide for a wide variety of student interests. Consequently all students can find problems to solve which match their ability and interests. Most students spend four days solving problems from the problem card deck. The teacher's guide provides suggestions for using the student booklet and the card deck, answers to all exercises, a record-keeping form, and two forms of a module test. The module test is usually administered on the tenth day and concludes the unit.

Throughout the instructional modules, students and their teachers are encouraged to use a four-step model in solving problems: (1) get to know the problem, (2) decide what to do, (3) do it, and (4) think back over what was done. This model was used in developing the booklets and the hints provided on the problem cards to help the students in finding solutions. It provides a language whereby students can communicate what they are doing and where they are having difficulty, as well as a general framework for attacking a problem.

A sample problem from the card deck in the Calculator I module is the following.

A ONE DOLLAR BILL, A TEN DOLLAR BILL, A 20 DOLLAR BILL, AND A 50 DOLLAR BILL EACH WEIGH ABOUT 1 GRAM. OF COURSE YOU WOULD RATHER HAVE 10 GRAMS OF \$10 BILLS THAN 10 GRAMS OF ONE DOLLAR BILLS. WHICH OF THESE BAGS WOULD YOU RATHER HAVE?



On the reverse side of the card containing this problem are the following hints, organized around the four-step model.

GET TO KNOW THE PROBLEM	What does one 20-dollar bill weigh? Will your answer be a number of grams, a number of bills, or one of the bags?
CHOOSE WHAT TO DO	How will you find the amount for bag A?
DO IT	Find the value of the money in each bag.
LOOK BACK	Did you find that both bags contained more than \$1000? Write a problem similar to this one.

Each instructional module was developed by IPSP using a loop design for product development. An eleven-step procedure was usually followed. Materials were: (1) written by IPSP personnel; (2) pilot-tested in classes taught by IPSP personnel in Price Laboratory School, Cedar Falls, Iowa, (3) evaluated; (4) revised; (5) pilot-tested in Price Laboratory School classes not taught by IPSP personnel; (6) evaluated; (7) revised; (8) field-tested in ten or more tryout schools; (9) evaluated; (10) revised, if necessary; and (11) prepared for dissemination.

The teachers involved in the tryout of IPSP materials were volunteers who had attended one of many half-day workshops conducted by IPSP personnel. The workshops acquainted the teachers with the four-step instructional model and with at least one of the eight IPSP modules. Feedback gathered from the tryout teachers was used in revising the materials. The workshops also provided a pool of approximately 350 Iowa teachers who volunteered to participate in the project. Most of the teachers were used to field test the materials while they were being developed and/or in the formal evaluation of IPSP conducted in 1978-79.

Once the modules had been sufficiently revised and successfully field tested, they were made available at cost to other educators interested in improving the problem-solving skills of students in grades 5-8.

Concurrently with the development of the instructional modules the IPSP team also developed a testing instrument. The broad goal of the IPSP test development was to produce a multiple-choice instrument which would measure problem-solving skills within each step of the four-step model.

The IPSP test was developed over a three-year period. By utilizing the Iowa Testing Program's tryout facilities it was possible to construct experimental units, administer them to representative samples of Iowa fifth through eighth graders and revise the units based on the item analyses and test data. The content validity was checked by a team of mathematics educators at several points in the development. Also, over 100 students were interviewed at various stages in the test development process as a concurrent check of test validity.

Like standardized tests the IPSP test can be efficiently administered to large groups of students and machine scored with various norm data easily obtainable. In addition, subtest scores corresponding to steps 1, 3, and 4 can be obtained. After nearly three years of effort, no viable way to test skills in step 2 in a multiple-choice format was found.

A brief description of these subtests and a sample item from each follows:

1. Step 1: Getting to Know the Problem

Items in this subtest require the student to identify extraneous or insufficient information in a problem setting, or to identify a question which could be answered using a given setting. Sample item:

> Joe bought 4 reflectors at $.50 each, a headlight for $2.98, a battery for $.35, and a roll of tape for $1.50. The clerk wanted to find the total cost. Which choice below would he need to know?
>
> 1) Joe had a free gift coupon for the battery.

2) The tape was a 2-inch roll.
3) The battery was a #3.
4) Joe paid with a credit card.

2. Step 3: Doing It

These items require a student to choose the correct computation needed to solve a problem, compute or estimate from a diagram, or apply a table or formula. Sample item:

> To convert a temperature reading from degrees Fahrenheit (F) to degrees Celsius (C), use this formula. $C = \frac{5}{9} \times (F - 32)$

What is 59^{o} Fahrenheit on the Celsius Scale?

1) 15^{o}
2) 18^{o}
3) 27^{o}
4) 74^{o}

3. Step 4: Looking Back

This subtest contains items which require the student to identify problems which can be solved in the same way as a given problem, determine the effect of varying the conditions in a given problem or determine if a given solution strategy is correct. Sample item:

Shelley has 75 marbles which is 11 more than twice as many as Karen has. To find how many marbles Karen has, Shelley added 75 + 11 and got 86. She then said Karen has 43 marbles. Is Shelley right?

1) Yes
2) No. She should have multiplied 86 x 2 and got 172.
3) No. She should have subtracted 75 - 11 = 64. Then 32 is the right answer.
4) No. She should have multiplied 11 x 2 = 22. Then 75 - 22 = 53 is the right answer.

The test was further validated by comparing IPSP scores with a) student scores in an interview-based test, and b) several subtests of the standardized Iowa Test of Basic Skills. Reliability analysis of the subtests and a test of the discrimination across subtests were also completed. The complete report of the validation procedures can be found in (Oehmke, 1979).

SUMMATIVE EVALUATION OF THE MATERIALS

In this section the results of the 1978-79 summative evaluation of the IPSP instructional modules are reported.

Procedure

Sample

During the first two years, the members of the IPSP team conducted inservice workshops for teachers throughout Iowa. Volunteers were sought from among the workshop participants to try the IPSP materials in their classrooms and to provide feedback to the writing team. This feedback was used to revise and improve the materials. In addition, teachers in the workshops were asked to volunteer to participate in the 1978-79 project evaluation.

In this way, a pool of over 200 fifth- through eighth-grade classrooms in Iowa was identified for the evaluation. For various reasons (teachers changing jobs, administrative difficulties, etc.), just 196 of these classrooms (4,708 students) were included in the evaluation at its outset. Fourteen of these classes did not complete the posttest, having dropped out of the study during the school year, again for various reasons. Thus, the results are based on the 182 classes (4,279 students) which were involved in the evaluation for its duration.

Instrumentation

A randomized pretest-posttest design with two experimental treatment groups and one control group was employed. The pretests and posttests which were administered are listed here. The attitude measures are all semantic differential scales.

1. IPSP Test (30 items)

   Step 1. Subtest- Getting to Know the Problem (10 items)

   Step 3. Subtest- Doing It (10 items)

   Step 4. Subtest- Looking Back (10 items)

   The pretest and posttests were essentially equivalent forms with one test used for 5th and 6th grades and a second test for the 7th and 8th grades.

2. WPATT: Attitude Toward Word Problems (6 items)
   Administered to all students both pretest and posttest.

3. CATT: Attitude Toward Calculators (6 items)
   Administered to all students both pretest and posttest.

4. TWPATT: Teachers' Attitude Toward Word Problems (15 items)
   Administered to all teachers both pretest and posttest.

5. TCATT: Teachers' Attitude Toward Calculators (19 items)
   Administered to all teachers both pretest and posttest.

6. MIS: Mathematics Instruction Summary (7 activity categories)
   Completed by all teachers once every two weeks during the evaluation to determine what was actually transpiring in the classrooms.

   The dates for the completion of the MIS were randomly chosen within each two-week time block with all teachers completing them on the same dates.

Treatments

Three treatment groups were defined as follows:

Treatment I: teachers taught two two-week IPSP modules during the study, one before Christmas and one after, using the teacher handbook, student booklets, problems decks and, if they were required in the modules, calculators.

Treatment II: teachers taught two two-week IPSP modules during the study, one before Christmas and one after, using only the problem decks and, if they were required in the modules, calculators.

Treatment III: teachers were given no special guidelines, materials or calculators, but were told to proceed as they ordinarily would.

Treatment I, then, is the full range of the IPSP approach with all materials and calculators (one for every two children) provided. Treatment II was included to isolate the effect of simply having more problems to solve and calculators to solve them, while Treatment III was the control, or "traditional" instruction, group.

Timetable

The schedule of events is outlined in Table 1.

(Insert Table 1 about here.)

Design

For purposes of most of the analysis, a 4 x 3 factorial design was employed with four levels of the fixed grade level effect and three levels of the random treatment effect. Classes within grades were randomly placed into one of the treatments. The class means were used as the statistical unit for analyzing all student test results, with the posttest means as dependent variables and the related pretest means as covariates. The design is illustrated in Figure 1. N is the number of classes completing the posttests in each cell.

(Insert Figure 1 about here.)

Hypotheses

The following null hypotheses were tested:

H1: There are no significant differences among treatment group posttest means (after adjustment for pretest differences) on the IPSP subtest 1, subtest 3, subtest 4 or total score.

H2: There are no significant differences among treatment group post-attitude means (after adjustment for pre-attitude differences) on the attitude toward word problem or attitude toward calculator scales.

H3: There are no significant differences among the post-attitude means of the teachers in the three treatment groups (after adjustment for pre-attitude differences) on the attitude toward word problems and attitude toward calculator scales.

H4: There are no significant differences among the mean times spent in each of the classroom activities listed on the MIS in the three treatment groups, and these means do not change significantly over observation time.

In addition to these four hypotheses, data were analyzed for grade level effects and grade x treatment interactions. A correlational analysis measuring the pairwise relationship between all variables was also completed.

Results

Test Reliabilities

Tests were kept as short as possible in order to allow the students to complete the pretesting and posttesting in one hour each. The reliabilities, which are a function of test length, were all within an acceptable range (10-item IPSP subtest reliabilities were typically about .70 with a .58 to .78 range, the 30-item IPSP test ranged from .81 to .87, the 6-item attitude scales ranged from .77 to .87, and the teacher attitude scales were over .90).

Tests of Hypotheses

Hypothesis One

The statistical test of H1 was carried out in several steps. Since grades 5 and 6 were given IPSP test forms which differed from those completed by grades 7 and 8, two separate analyses were done with grades 5 and 6 together and grades 7 and 8 together, respectively. For each analysis a 2 x 3 mixed factorial design, two levels of the fixed grade effect and three levels of the random treatment effect, was employed. The statistical unit was classroom means, since it can hardly be argued that individual students are independent of one another in the design. The design allows us to consider grade x treatment interactions as well as grade and treatment main effects. The two designs are diagrammed in Figure 2. N is the number of classes which completed the posttest in each cell.

(Insert Figure 2 about here.)

Associated with each class were six means, one for each subtest for both the pretest and the posttest. This called for a multivariate analysis.

In particular, a multivariate analysis of covariance (MANCOVA) was run for each design using the three pretest means as covariates and the three posttest means as dependent variables. Logically, the MANCOVA was run at the .05 level of significance. If the MANCOVA F-test failed to reach the .05 level, this indicated that the differences in the adjusted means for each cell in the design could have been due to chance as the samples were drawn randomly from the same population, and no further test was run. However, if the MANCOVA resulted in an F-value that was beyond the critical F at the .05 level, univariate ANCOVA's using the individual dependent variables were run to determine if any individual variable was contributing to the differences detected by the MANCOVA. Finally, if no ANCOVA yielded a significant F no further analysis was made, but if one or more ANCOVA yielded a significant F, post hoc pairwise two-tailed t-tests were used to determine which cell means differ significantly. In fact, it is a bit more complicated than this because significant interactions sometimes make the interpretation of main effects difficult.

In practice, the univariate tests are often run if the MANCOVA F comes close to reaching the critical value. However, in that case the ANCOVA results should be considered to be somewhat tentative.

Adjusted posttest means including total IPSP test means for each cell for grades 5 and 6 are given in Table 2. The MANCOVA F-values for overall grade x treatment interaction ($F(6,158) = 0.96$; $p < .46$), overall grade effect ($F(3,80) = 0.60$; $p < .62$) and overall treatment effect ($F(6,158) = 1.62$; $p < .14$) did not reach the .05 level of significance.

However, since the treatment F was "fairly" close to significance the univariate ANCOVA's were run. The results are given in Table 3.

(Insert Tables 2 and 3 about here.)

The mean for treatment I was highest on all three subtests but the F, in each case, failed to reach the .05 probability level. However, the treatment F for the total IPSP test did reach the .06 level, rather strong evidence of a treatment effect. A follow-up t-test showed that the group I mean was greater than the group III mean ($t = 2.42$, $p < .05$) but no other pairwise contrasts were significant.

Adjusted posttest means for grades 7 and 8 are given in Table 4. The MANCOVA F-values for overall grade x treatment interaction ($F(6,142) = 0.47$; $p < 0.83$), overall grade effect ($F(3,72) = 0.40$; $p < 0.75$) and overall treatment effect ($F(6,142) = 1.01$; $p < 0.42$) did not reach the .05 level of significance. There was no clear trend of superiority for one treatment over another in grades 7 and 8, although treatment II means are slightly lower than those of the other two treatments.

(Insert Table 4 about here.)

Hypothesis Two

Since students in all four grades completed the same attitude scales, all grade levels were included in the same design, namely the 4 x 3 factorial design shown in Figure 1. The statistical analysis to test H2 followed the same pattern as the achievement analysis. Responses on the semantic differential scales were assigned values from 1 to 5; thus possible scale scores ranged from 6 to 30; with a high score indicating a positive attitude. The adjusted post-attitude means are given in

Table 5. For the MANCOVA, the two pre-attitude scales were the co-variates, and the two post-attitude scales were the dependent variables. The MANCOVA yielded a significant overall grade x treatment F ($F(12,326) = 2.12$; $p < 0.02$), a significant F for overall grade effect ($F(6,326) = 4.02$; $p < 0.001$) and a significant F for overall treatment effect ($F(4,326) = 3.78$; $p < .01$).

(Insert Table 5 about here.)

Consequently, univariate ANCOVA's were run for each of the attitude scales. The results of the ANVOCAs using the two attitude measures separately as dependent variables are given in Table 6.

(Insert Table 6 about here.)

Since the grade x treatment interactions are significant or nearly so in both cases, the adjusted cell means were examined. These are plotted in Figure 3.

(Insert Figure 3 about here.)

From the plots it appears that attitude toward word problems in treatment I was generally higher than in the other two treatments. In fact, this is the case based on the t-tests. However, the attitude toward calculator plot is less clear-cut. There also appears to be a steady decline in both attitude scales as grade level increases. The results of the pairwise comparisons using follow-up t-tests are given in Table 7.

(Insert Table 7 about here.)

Hypothesis Three

The design used to test H3 was the same as that used for the student attitudes as illustrated in Figure 1. The covariates for the MANCOVA were the two pre-attitude scores of the teachers and the dependent variables were the post-attitude scores. In this case, the statistical unit was the individual teacher's score.

The statistical analysis followed the same pattern as the analyses for student achievement and attitudes. The adjusted post-attitude means are given in Table 8. The MANCOVA resulted in a significant overall treatment effect ($F(4,276) = 2.45$; $p < .05$), but the overall grade effect ($F(6,276 = 0.36$; $p < .91$) and the overall grade x treatment interaction ($F(12,276) = 0.52$; $p < 0.90$) were not significant. Consequently, univariate ANCOVA's were run with scores on each attitude scale as dependent variable to isolate the treatment effect. The results of the univariate ANVOCA's are given in Table 9.

(Insert Tables 8 and 9 about here.)

The only significant F is the treatment effect for the attitude toward calculators scale. The follow-up t-test shows that the treatment I mean is greater than the treatment III mean ($t = 3.02$, $p < .01$) but the other differences are not statistically significant.

Hypothesis Four

In order to ensure that teachers were following the directions for their treatment, to examine possible effects of the treatment on teacher behavior, especially over time, and to further describe the classrooms and treatments, the teachers were asked to complete the MIS ten times during the study. This provided measures of the teacher's perception of the number of minutes spent that day on each of seven instructional activities and the number of students absent. The instructional activities are: 1) teaching or working on word problems other than IPSP materials (ACT 1), 2) teaching or working on IPSP materials (ACT 2), 3) use of calculator by one or more students (ACT 3), 4) teacher-led presentation to large group, i.e., 8 or more students (ACT 4), 5) small group work, i.e., 2 to 7 students (ACT 5), 6) individual work (ACT 6), and 7) mathematics testing (ACT 7).

Teachers checked a range of minutes for each instructional activity, i.e., 1-9, 10-19, 20-29, 30-39, 40-49, 50-59 and 60 or more. For purposes of analysis, each check mark was assigned the value of the midpoint of that interval. The last choice, 60 or more, was counted as 60. Grade by treatment cell means averaged over the ten observation times are given in Table 10.

(Insert Table 10 about here.)

Seven repeated measures ANOVA's were run each with one of the seven instructional activity variables as dependent variable. Only those teachers who returned all ten of the MIS forms were considered in this analysis. In order to balance the design, some classrooms were randomly removed where cell numbers were not in proportion. A 4 x 3 x 10 (grade x treatment x observation time) design was employed. The cell sizes are shown in Figure 4.

(Insert Figure 4 about here.)

The repeated measures ANOVA's resulted in no interpretable time effect or time x treatment interactions. It was concluded that the instructional activity patterns did not change systematically over time, and this pattern was not differentially affected by the treatments. In the remainder of the analysis, classroom mean times for each instructional activity across the ten observations are the dependent variables.

Seven 4 x 3 univariate ANOVA's were run with each instructional activity as the respective dependent variable. To save space the results are not reported here. However, there is ample evidence that treatments I and II actually took place as defined. The four weeks planned for IPSP instruction within the 20-week study compares well with 19.2% and 17.1% reported by the teachers for IPSP activities in groups I and II, respectively. In addition, about 17% of class time involved calculator use in groups I and II. On the other hand, reported time using IPSP materials was virtually zero in the control group; and although the control teachers were not restricted from using the calculator by definition, only about 2% of their instructional time involved a calculator.

There were also rather consistent, statistically significant differences in the uses of small group and teacher-led presentation between grade levels. Generally, the higher the grade level the less the reported use of small group instruction and the more time spent in teacher-led presentations.

Correlational Findings

A Pearson Product Moment Correlation matrix was computed giving measures of the pairwise relationships between all student and teacher variables. Again, for all student variables the class mean is the statistical unit. Each correlation coefficient was computed using all classes or teachers for which data was available for the two variables under consideration. Statistically significant relationships between pretest measures only which appear to have educational significance are reported here. It was felt that relationships involving post-measures would be confounded by treatment effects.

1. Of the three IPSP subtests, step 1 and step 4 are most closely related ($r = .91$). This finding is consistent with test validation results reported elsewhere (Oehmke, 1979). In fact, there appears to be little difference between the underlying abilities tested by the Step 1 and Step 4 subtests. However, Step 3 ($r = .76$ with Step 1 and $r = .82$ with Step 4) is somewhat different from the other two. We hypothesize that the Step 1 and Step 4 subtests both require a good deal of verbal reasoning ability, while the Step 3 appears to have a heavier requirement of quantitative skill.
2. The teacher's attitude toward problem solving was related positively to the problem-solving ability of the class ($r = .18$, $p < .05$), while the amount of time the teacher reported spending on individual work was related negatively ($r = -.20$, $p < .05$), to the problem-solving ability of the class.
3. There was a positive, though small, correlation between the class's attitude toward word problems and their problem-solving ability but only on the Step 3 test($r = .15$, $p < .05$). This positive finding is a result seen in many previous studies. The absence of a positive correlation between attitude and the other subtests is interesting, and may reflect the difference between Step 3 and the other two subtests.

Discussion

This study involved over 4,000 eighth grade students in a cross-section of Iowa Classrooms. The large sample size made it possible to use class means as the statistical unit. Teacher self-report data indicate that the two experimental treatments were actually used during the study. Since the testing instruments also were quite reliable, there is good reason to consider the results of this study to be stable and accurate as well as generalizable, at least to other Iowa classrooms.

The effectiveness of the IPSP materials and approach was illustrated in several ways. First, there was a superiority ($p < .06$) of the IPSP treatment group over the control group on the total IPSP test in grades 5 and 6. Second, classes getting the IPSP treatment exhibited significantly more positive attitudes toward word problems. Third, classes getting the IPSP treatment exhibited significantly more positive attitudes toward calculators than the problem deck only group. Fourth, teachers who used IPSP materials compared to those who did not exhibited significantly more positive attitudes toward calculators after the study. These results when taken together are strong evidence that the IPSP approach is effective - more effective than just giving students more problems to solve and more effective than the "usual" approaches taken by teacher. A 1979-80 follow-up evaluation which examined the effects on several intact classes of a second year of IPSP was also completed. At the time of this writing these data have not been analyzed.

There are several findings in this study which are unrelated to the effectiveness of the IPSP materials, but which may be of interest to other researchers and teachers. Some of these results are further evidence to support previous findings. First, attitudes of students, both toward word problems and toward calculators, became less positive with the age of the student, but teacher attitudes did not differ significantly across the four grade levels. Second, the relationship between student attitude toward word problems and achievement scores was positive but small. Third, teacher-led presentation is used more at the higher grades than at the lower grades while the reverse is true for small group instruction. Fourth, the attitude of the teacher toward word problems was positively related to the class's performance. Fifth, the amount of time spent in individual instruction is negatively related to the class's problem -solving achievement.

The last two correlational findings are the most interesting. The fourth finding suggests, though no causation is proven, the teacher's attitude does affect the students' performance in problem solving. The fifth finding can be interpreted in the light of two different areas of recent research. First, it is possible that time spent in individual seat work is often not profitably spent; hence, negatively affecting student performance. Recent studies showing on-task time to be a crucial variable in the classroom are consistent with this finding (Berliner, 1978). Second, the negative relationship between time spent in individual instruction and the class's problem-solving performance may be explained in terms of research on individualized or self-paced instruction. Since, teachers who used more individual instruction also reported more testing ($r = .21$, $p < .05$), it is likely that some of these classrooms were using a form of self-paced instruction which involves much individual work and frequent testing. Recent research suggests that this approach often results in poor student achievement scores (Schoen, 1977).

FURTHER DISCUSSION

The results of the evaluation reported here attest to the effectiveness of the IPSP approach and materials as measured by posttesting, and by changes in attitudes of students and teachers. How do these findings fit into the broader context of research on teaching problem solving?

In an area of research which has often been written about in tones of despair, the findings in favor of the IPSP treatments seem especially impressive. In fact, they are consistent with the findings in a recent meta-analysis of this research. Marcucci (1980) examined 33 studies and found that, of four types of methods of teaching problem solving, heuristic methods have had the most positive effect on student learning, though positive effects have not always been shown. More research and development in the teaching of problem solving are certainly needed, but there is mounting evidence that, as Carpenter et al (1980) and others have suggested, the most promising direction is via the teaching of specific heuristics.

REFERENCES

Berliner, David C., Allocated time, engaged time, and academic learning time in elementary school mathematics. Paper presented at the 56th Annual Meeting of the National Council of Teachers of Mathematics, San Diego, April 1978.

Callahan, Leroy G., & Glennon, Vincent J. How can we improve ability to solve verbal problems? In Elementary School Mathematics, A Guide to Current Research (4th ed.). Washington, D.C.: Association for Supervision and Curriculum Development, 1975.

Carpenter, Thomas P., Coburn, Terrence G., Reys, Robert E., & Wilson, James W. Notes from national assessment: Word problems. The Arithmetic Teacher, 1976, 23.

Carpenter, Thomas P., Corbitt, Mary K., Kepner, Jr., Henry S., Lindquist, Mary M., & Reys, Robert E. NAEP note: Problem solving. The Mathematics Teacher, 1980, 73, 427-433.

Lester, Frank K. Research in mathematical problem solving. In Richard J. Shumway (Ed.), Research in Mathematics Education. Reston, Virginia: National Council of Teachers of Mathematics, 1980.

Marcucci, Robert G. A meta-analysis of research on methods of teaching mathematical problem solving. Unpublished doctoral dissertation, University of Iowa, 1980.

National Advisory Committee on Mathematical Education. Overview and analysis of school methematics, grades K-12. Washington, D.C.: Conference Board of Mathematical Sciences, 1975.

Oehmke, Theresa M. The development and validation of a testing instrument to measure problem solving skills of children in grades five through eight. Unpublished doctoral dissertation, University of Iowa, 1979.

Schoen, Harold L. Implications of research for instruction in self-paced mathematics classrooms. In F. Joe Crosswhite & Robert E. Reys (Eds.), Organizing for Mathematics Instruction. Reston, Virginia: National Council of Teachers of Mathematics, 1977.

Table 1

Time Table

Date	Activity
9/15/78	Pretests administered
9/26/78	First MIS completed
10/12/78	Second MIS completed
10/16/78	Third MIS completed
10/30/78	Fourth MIS completed
11/16/78	Fifth MIS completed
12/06/78	Sixth MIS completed
9/15-12/21/78	Treatments I and II teach first two-week IPSP module
1/16/79	Seventh MIS completed
1/29/79	Eighth MIS completed
2/06/79	Ninth MIS completed
2/20/79	Tenth MIS completed
1/3-3/2/79	Treatments I and II teach second two-week IPSP module
3/03/79	Posttest administered

Table 2

Adjusted Posttest Means

Achievement-Grades 5 and 6

Treatment	Grade	Step 1	Step 3	Step 4	Total
I	5	6.84	7.25	5.84	20.06
I	6	7.04	7.47	6.27	20.71
I OVERALL		6.94	7.36	6.05	20.38
II	5	6.82	7.26	6.04	20.16
II	6	6.89	7.30	5.74	20.17
II OVERALL		6.86	7.28	5.89	20.16
III	5	6.81	7.37	5.72	19.77
III	6	6.61	7.19	5.75	19.64
III OVERALL		6.71	7.28	5.74	19.71

Table 3

ANCOVA Results

Achievement-Grades 5 and 6

	Source	DF	SS	F	Prob.
S1	Grade	1	0.07	0.29	0.59
	Treatment	2	0.61	1.35	0.26
	G x T	2	0.66	1.46	0.24
S3	Grade	1	0.00	0.00	0.95
	Treatment	2	0.19	0.65	0.53
	G x T	2	0.18	0.59	0.56
S4	Grade	1	0.32	1.41	0.24
	Treatment	2	0.66	1.41	0.25
	G x T	2	0.65	1.42	0.25
Total	Grade	1	0.55	0.43	0.52
	Treatment	2	7.62	2.98	0.06
	G x T	2	2.84	1.11	0.33

Table 4

Adjusted Posttest Means

Achievement-Grades 7 and 8

Treatment	Grade	S1	S3	S4	Total
I	7	6.74	6.40	5.97	19.11
I	8	6.58	6.36	5.89	18.82
I OVERALL		6.66	6.38	5.93	18.97
II	7	6.53	6.33	5.90	18.74
II	8	6.39	6.09	5.58	18.06
II OVERALL		6.46	6.21	5.74	18.40
III	7	6.58	6.38	6.01	18.96
III	8	6.44	6.48	6.01	18.95
III OVERALL		6.51	6.43	6.01	18.96

Table 5

Adjusted Posttest Means

Student Attitude

Treatment	Grade	WPATT	CATT
I	5	18.81	21.01
I	6	18.64	20.67
I	7	18.28	20.91
I	8	18.14	20.33
I OVERALL		18.47	20.73
II	5	18.19	20.57
II	6	18.42	20.82
II	7	17.95	20.28
II	8	18.17	20.33
II OVERALL		18.18	20.50
III	5	18.66	20.92
III	6	17.92	20.91
III	7	18.26	20.59
III	8	18.04	20.72
III OVERALL		18.22	20.79
Grade Overall	5	18.55	20.83
	6	18.33	20.80
	7	18.16	20.60
	8	18.11	20.46

Table 6

ANCOVA Results

Student Attitudes

Scale	Source	DF	SS	F	Prob.
WPATT	Grade	3	5.01	4.66	0.004
	Treatment	2	2.67	3.73	0.03
	Grade x Treatment	6	5.55	2.58	0.02
CATT	Grade	3	3.95	4.12	0.01
	Treatment	2	2.67	4.17	0.02
	Grade x Treatment	6	3.35	1.75	0.11

Table 7

Pairwise Comparisons

Student Attitude

Overall Effects

Scale	Treatment			Grade		
	Result	t	p	Result	t	p
Attitude WPATT	I > II	2.64	<.01	5 > 8	3.67	<.01
	I > III	2.27	<.05	5 > 7	3.00	<.01
Attitude CATT	I > II	2.30	<.05	5 > 8	3.08	<.01
	III > II	2.90	<.01			

Grade x Treatment Effects

Scale	Grade 5			Grade 6			Grade 7			Grade 8		
	Result	t	p	Result	t	p	Result	t	p	Result	t	p
Attitude WPATT	I > II	2.95	<.01	I > III	3.43	<.01	NSD			NSD		
	III > II	2.14	<.05	II > III	2.38	<.05						
Attitude CATT	I > II	2.20	<.05	NSD			I > II	2.63	<.02	NSD		

Table 8

Adjusted Posttest Means

Teacher Attitude

Treatment	Grade	TWPATT	TCATT
I	5	48.24	61.65
I	6	46.85	61.80
I	7	47.23	61.00
I	8	46.75	61.08
I OVERALL		47.27	61.38
II	5	46.06	60.31
II	6	46.82	61.11
II	7	45.90	59.60
II	8	46.61	61.42
II OVERALL		46.35	60.61
III	5	46.16	60.26
III	6	47.59	59.33
III	7	47.45	58.77
III	8	46.85	59.66
III OVERALL		47.01	59.51

Table 9

ANCOVA Results

Teacher Attitude

Scale	Source	DF	SS	F	Prob.
TWPATT	Grade	2	2.45	0.09	0.96
	Treatment	3	21.59	1.17	0.31
	Grade x Treatment	6	40.52	0.73	0.62
TCATT	Grade	2	21.50	0.61	0.62
	Treatment	3	87.21	3.70	0.03
	Grade x Treatment	6	21.47	0.30	0.93

Treatments

GRADE	I	II	III
5	N=19	N=15	N=16
6	N=17	N=16	N=16
7	N=11	N=12	N=15
8	N=13	N=16	N=16

Figure 1. Research Design

Treatments

GRADE	I	II	III
5	N=19	N=15	N=16
6	N=17	N=16	N=16

Treatments

GRADE	I	II	III
5	N=11	N=12	N=15
6	N=13	N=16	N=16

Figure 2. Research Design for Achievement Criteria

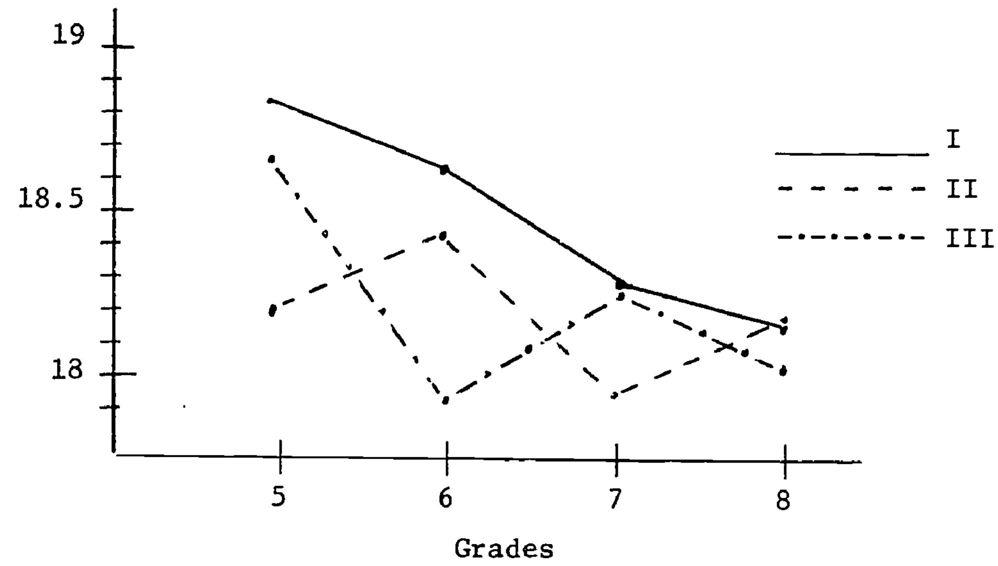
TABLE 10

Instructional Activity Means in Minutes Per Day

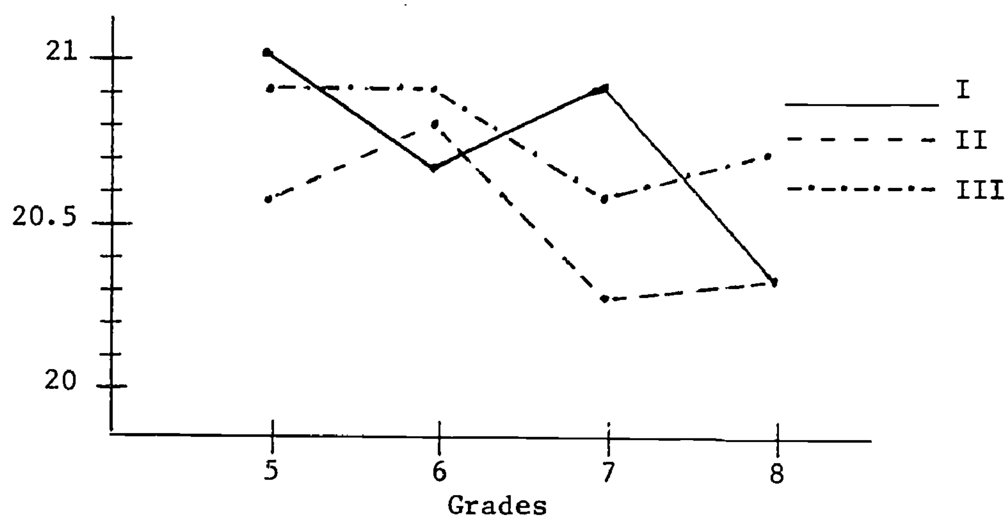
Variable	Treatment	Grade 5	Grade 6	Grade 7	Grade 8	Treatment Mean	Percent of Time/45 min.
ACT 1	I	8.83	4.56	6.56	6.39	6.59	14.6
	II	5.40	3.95	6.60	5.35	5.33	11.8
	III	6.69	4.50	3.38	2.94	4.38	9.7
	Grade Mean	6.93	4.32	5.63	4.98	5.47	12.2
ACT 2	I	6.89	8.72	7.33	11.61	8.64	19.2
	II	8.25	7.45	9.80	5.30	7.70	17.1
	III	0.63	0.00	0.13	0.00	0.19	0.4
	Grade Mean	5.54	5.67	6.11	5.83	5.79	12.9
ACT 3	I	5.78	7.28	7.89	9.11	7.52	16.7
	II	9.66	7.85	8.70	5.27	7.87	17.5
	III	0.81	0.81	0.63	1.20	0.86	1.9
	Grade Mean	5.74	5.30	6.04	5.34	5.61	12.5
ACT 4	I	10.00	9.39	12.17	16.22	11.95	26.6
	II	11.47	13.80	11.00	15.75	13.01	28.9
	III	13.38	18.69	10.38	17.94	15.10	33.6
	Grade Mean	11.55	13.78	11.21	17.22	13.44	29.9
ACT 5	I	12.72	10.43	8.89	12.78	11.21	24.9
	II	13.35	6.05	9.35	5.65	8.60	19.1
	III	11.06	4.81	3.25	4.90	6.01	13.4
	Grade Mean	12.46	7.14	7.39	7.80	8.70	19.3
ACT 6	I	18.11	15.72	16.89	17.33	17.01	37.8
	II	20.05	13.75	20.55	13.65	17.00	37.8
	III	20.88	18.63	9.88	17.51	16.73	37.2
	Grade Mean	19.65	15.85	16.17	16.02	16.92	37.6
ACT 7	I	5.39	6.78	6.28	4.00	5.61	12.5
	II	8.89	4.15	7.05	2.60	5.67	12.6
	III	4.64	4.75	6.25	4.69	5.08	11.3
	Grade Mean	6.46	5.20	6.56	3.69	5.48	12.2
Absentees	I	1.24	0.89	1.39	0.98	1.13*	4.8**
	II	1.17	0.86	1.16	1.30	1.12	4.8
	III	0.95	0.91	1.59	1.46	1.23	5.2
	Grade Mean	1.13	0.88	1.36	1.24	1.15	4.9

ber of student's absent per day.

cent of students based on an overall class average of 23.5.



WPATT



CATT

Figure 3. Plot of Adjusted Attitude Means

GRADE	Treatment I	Treatment II	Treatment III
5	N=9	N=10	N=8
6	N=9	N=10	N=8
7	N=9	N=10	N=8
8	N=9	N=10	N=8

Figure 4. Research Design for Instructional Activities